182

New Method for Interpreting Data Obtained in Service Tests on Sole Leather**

By C. W. Mann

*Eastern Regional Research Laboratory**
Philadelphia, Pennsylvania

Introduction

An accurate appraisal of resistance to wear is essential in tests on the wearing quality of sole leather. Correct interpretation is of major importance. Otherwise the results may lead to erroneous conclusions and lack of confidence in wear tests in spite of the fact that the testing method is satisfactory. Since the results of abrasion tests thus far developed show poor correlation with actual wear, service tests are the only reliable means of evaluating wear resistance.

If only two leathers are to be compared, the problem is simplified somewhat by the possibility of equalizing a large part of the variation due to wearers and wearing conditions. Frey, Clarke and Leinbach[1] compared vegetable tanned and chrome retanned sole leathers by using a sole from each tannage on every pair of shoes in the test. The relative wear resistance was calculated as the average ratio of the thickness losses in the various pairs. Whitmore and Downing[2] describe a method for comparing two leathers on the same pairs of shoes. In this method the relative rating is the geometric mean of the ratios of thickness losses. Either of these methods may be applied to data for which the length of time the shoes were worn is unknown.

A search of the literature failed to disclose a satisfactory method for solving the more general problem of interpreting results of experiments in which a group of leathers are compared. The usual method is to compare the rates of wear, such as the number of hours required to wear through a given thickness. The rate of wear is subject to relatively large errors because of the variation due to wearers and wearing conditions as well as the difficulty in obtaining reliable and consistent records of wearing time. This method requires that conditions be controlled to minimize causes of variation, such as mistreatment of shoes, physical defects of wearers, occupational factors, and unusual kinds of exposure. As this close supervision and control would require a greater expenditure of time than is generally available and better cooperation than wearers are usually willing to give, the method must be considered impractical.

In their recent "Camp Lee Tests" Hobbs and Kronstadt[3] compared the wearing quality of 20 commercial tannages. In this experiment comparisons

** Presented at the Eastern Regional Meeting of the American Leather Chemists Association, New York, N. Y., October 31, 1945.

* One of the laboratories of the Bureau of Agricultural and Industrial Chemistry, Agricultural Research Administration, United States Department of Agriculture.

between the different tannages were obtained by making up each pair of taps with a sole from one tannage matched against a sole from another tannage, as it was not practicable for the men at Camp Lee to keep daily records of the number of hours that the shoes were worn. The results were calculated by the formula $Qx = \frac{2y}{x + y}$, where Qx is the "quality ratio" of a tap X in a pair, XY, and x and y are the thicknesses worn away from X and Y, respectively. The average quality ratio for each tannage was calculated, and the leathers were compared on this basis.

A study of this method of calculation has shown that the average quality ratio underestimates the magnitude of the differences between tannages, although it probably evaluates the significance of differences adequately. Thus, while the method may have been satisfactory for the purpose for which it was used, it might lead to serious errors elsewhere, as for example in estimating the economic value of an impregnating process. The purpose of this paper is to describe an improved method of interpreting the results of experiments of this type. This method gives more uniform and accurate results.

Development of the Method

Service tests of sole leather involving more than two tannages require that the soles be worn under various conditions. The relative wear resistance of the tannages will be affected by the following causes of variations:

1. Differences in soles from the same tannage.
2. Wearing conditions—wearing time, weather, soil, etc.
3. Differences in wearers.
4. Differences between right and left shoes of the same wearer.

By considering each pair as a separate test, and determining the relative wear resistance of each sole in the pair, a large part of the variation is eliminated. Since each pair is tested by the same wearer under the same conditions, these two sources of variation do not affect the results for the pair. The remaining variation may be considered as experimental error.

It would be desirable to have a standard soling material with which any leather might be compared in a service test. Since no satisfactory material has yet been found for this purpose, each experiment must have a separate standard, which furnishes the relative ratings of the leathers in that experiment but cannot be used for comparison with leathers in other tests. A convenient and reliable standard with which a group of leathers may be compared is the average for the entire group.

The comparison of the wear resistance of a given tannage with the average for a group is made by testing that tannage on pairs of shoes with each of the other tannages in the group and calculating the average ratio of the thickness losses of the soles in the pairs. Since the average of a series of ratios is given by the geometric mean, to compare leather A with the series of leathers X_1, X_2, etc., the formula

$$R_A = \left(\frac{x_1}{a_1} \cdot \frac{x_2}{a_2} \cdot \frac{x_3}{a_3} \cdots \frac{x_n}{a_n}\right)^{\frac{1}{n}} \qquad (1)$$

may be used, in which x_1, x_2, etc., are the thickness losses in service of soles from the corresponding tannages, a_1, a_2, etc., are losses from tannage A, and n is the number of pairs. R_A, the "wear rating" of tannage A, is the geometric mean of the series of ratios of thickness losses of pairs in which soles from tannage A were worn.

If the leathers X_1, X_2, etc., and leather A are considered separately, formula (1) may be written $R_A = \frac{x_m}{a_m}$, in which x_m and a_m are the geometric mean losses of the leathers X_1, X_2, etc., and leather A, respectively. Thus the method compares the mean loss of a given leather with the mean loss of a series of leathers worn on the same pairs of shoes. Any number of leathers may be compared with leather A by testing each with the same series of leathers, X_1, X_2, etc. A convenient form of equation (1) is obtained by converting to logarithms. Thus

$$\log R_A = \frac{1}{n}\left[\log\left(\frac{x_1}{a_1}\right) + \log\left(\frac{x_2}{a_2}\right) + \log\left(\frac{x_3}{a_3}\right) \ldots + \log\left(\frac{x_n}{a_n}\right)\right]$$

The application of the method is best explained by an example.

Application to the Camp Lee Data

The data obtained at Camp Lee will be used to demonstrate the application of the method. As an example of the procedure used in the calculations, the data for tannage A are given in Table I, together with values necessary to obtain the geometric mean wear rating, R_A.

As this method involves the calculation of geometric mean ratios, it is convenient to tabulate the logarithms of the ratios of thickness losses for each pair of soles as shown in Table I. The ratios of thickness losses are multiplied by 100 to convert them to a percentage basis; this assigns the value 100 to the mean ratio for the entire group of comparisons. In order to avoid errors arising from the variation in the number of replications of the different comparisons, equal weight is given to the results for each comparison of two tannages by calculating the geometric mean rating from the averages of replicate ratios.

In the method of inter-comparison used in this experiment, each leather is compared with every leather in the group, therefore each leather rating must include the average ratio (1.00) obtained in a comparison with itself. Failure to include this value introduces errors of approximately 1 per cent in the highest and lowest ratings of a group of 20 tannages. Although such errors may not be considered important in an experiment of this size, they

TABLE I

THICKNESS LOSSES IN SERVICE OF TANNAGE A AS COMPARED WITH THOSE OF OTHER TANNAGES IN THE CAMP LEE TEST; CALCULATION OF THE WEAR RATING, R_A

Tannage of Mate	Thickness Loss in Service: a 1/1000 in.	Thickness Loss in Service: Mate (X) 1/1000 in.	Ratio of Thickness Losses $100 \frac{x}{a}$	$\log\left(100 \frac{x}{a}\right)$	Average
A	*	...	100.0	....	2.000
B	142	155	109.2	2.038	2.038
C	143	144	100.7	2.003	
	150	171	114.0	2.057	
	210	152	72.4	1.860	1.973
D	68	71	104.4	2.019	
	163	175	107.4	2.031	
	113	118	104.4	2.019	2.023
E	149	105	70.5	1.848	
	104	90	86.5	1.937	
	128	103	80.5	1.906	1.897
F	166	126	75.9	1.880	
	101	47	46.5	1.668	
	97	73	75.3	1.877	1.808
G	118	106	89.8	1.953	
	145	152	104.8	2.020	
	161	150	93.2	1.969	1.981
H	103	109	105.8	2.024	
	172	192	111.6	2.048	2.036
J	211	147	69.7	1.843	
	149	128	85.9	1.934	
	141	122	86.5	1.937	1.905
K	151	116	76.8	1.885	
	171	123	71.9	1.857	
	171	147	86.0	1.934	1.892
L	177	153	86.4	1.936	
	120	80	66.7	1.824	
	194	214	110.3	2.042	1.934
M	62	86	138.7	2.142	
	187	190	101.6	2.007	
	165	180	109.1	2.038	2.062
N	176	162	92.0	1.964	
	188	169	89.9	1.954	1.959
P	180	159	88.3	1.946	
	152	211	138.8	2.142	
	131	117	89.3	1.951	2.013
Q	146	129	88.4	1.946	
	92	66	71.7	1.856	
	142	137	96.5	1.984	1.929
R	66	70	106.1	2.026	
	126	166	131.7	2.120	2.073
S	85	93	109.4	2.039	
	194	184	94.8	1.977	2.008
T	150	115	76.7	1.885	
	147	128	87.1	1.940	
	175	157	89.7	1.953	1.926
U	65	56	86.2	1.936	
	155	144	92.9	1.968	1.952
W	179	183	102.2	2.009	
	133	105	78.9	1.897	
	199	144	72.4	1.860	1.922
Grand Average..........					1.9666
Geometric mean wear rating, R_A..........					92.6

*In duplicate comparisons of tannage A with itself the losses (113–124) and (174–167) were obtained. The geometric mean of the 4 ratios,

$\frac{113}{124}, \frac{124}{113}, \frac{174}{167}$, and $\frac{167}{174}$, is 1.00.

become more important in smaller tests because of the inverse relationship between the size of the error and the number of tannages compared. The ratings of the other tannages were calculated in the same manner as that of tannage A.

Discussion

In Table II the original quality ratios and the wear ratings calculated by the new method are compared. The tannages are arranged in the order of rank on the basis of wear ratings. The close agreement between results of the methods in the order of rank indicates that the original method rates the leathers in the correct order. However, the range of the numerical ratings of the quality ratios is only about six tenths that obtained by the new method. Quality ratios indicate that the most resistant leather wore 16.5 per cent $\left(\frac{109.0}{93.5}\right)$ better than the least resistant leather, whereas wear ratings indicate that this difference is 28.1 per cent $\left(\frac{116.7}{91.1}\right)$. The smaller range of the quality ratios is due to a fault of the original method which results in a shift of high and low ratings toward the general mean. This shift is easily demonstrated by means of data in which the wearing time is known or with hypothetical data.

An analysis of variance[4] of the logarithms of thickness loss ratio shows that the differences between the tannages are highly significant, as compared

TABLE II

COMPARISON OF WEAR RATINGS AND QUALITY RATIOS OF TANNAGES IN THE CAMP LEE TEST

Tannage	Rank	Wear Rating	Rank	Quality Ratio
F	1	116.7	1	109.0
L	2	115.7	2	107.7
K	3	113.0	3	107.2
Q	4	108.4	4	104.8
P	5	105.4	5	101.8
J	6	104.5	6	101.5
W	7½	101.9	9	99.8
B	7½	101.9	10½	99.4
M	9	101.1	13	97.8
G	10	100.1	7	102.4
E	11	97.7	10½	99.4
T	12	97.1	8	100.3
H	13	95.0	14½	97.2
S	14	94.1	16	95.9
D	15	92.9	12	98.2
A	16	92.6	19	95.3
R	17	92.4	17	95.7
C	18	92.3	18	95.4
U	19	92.0	14½	97.2
N	20	91.1	20	93.5

with the error variance, calculated from the deviations of replicate values from their means. Thus the ratio of variances, $F = 7.03$, is based upon 19 and 219 degrees of freedom. From the standard deviation, 0.0795, it was possible to calculate the minimum difference between two means which may be considered significant. This logarithmic value is 0.0349; therefore the minimum significant difference is 8.37 per cent of the lower mean.

The analysis showed that 4 tannages (F, L, K, and Q) were better than average and two (N and U) worse. Although 4 tannages (A, C, D, and R) just failed to differ significantly from the average, they were less resistant to wear than the 9 top-ranking tannages.

In the usual service test, a few pairs of shoes are only slightly worn. Soles from these shoes, worn just through the grain surface, may give less reliable results than those worn for a longer time. It is generally agreed that low values of thickness losses should be rejected, but no definite value has been selected as the dividing line. In studying this question with the Camp Lee data, comparisons (of two tannages) in which the minimum loss was less than 40 mils were examined. The deviations of replicate values (logarithms) from their means were highly significant, as compared with error. ($F = 3.04$, for 9 and 219 degrees of freedom; 1 per cent point, 2.50). Similar calculations for the ranges 40 to 50 mils and 50 to 60 mils showed that deviations in these groups were not significantly different from error. This analysis indicates that losses in the range below 40 mils are more variable and therefore less reliable than higher values, whereas losses in the ranges 40 to 50 mils and 50 to 60 mils are no more variable than the average for the entire group. On this basis, 40 mils was selected as the minimum loss in thickness to be included in the data.

Accidents resulting in loss of data may leave one or more tannages with no direct comparison. Omitting a value may introduce a relatively large error if the two tannages differ in resistance to wear; therefore an adjustment should be made by estimating the missing value. A convenient and adequate estimate is obtained as the ratio of the ratings of the two leathers based on the available data. After this ratio is calculated, it is inserted in the table as the missing thickness loss ratio, and the corrected geometric mean rating is recalculated.

The variation in replicate ratios of thickness losses in column 4 of Table I shows the weakness of conclusions based upon a single comparison or even a small number of replicate comparisons. For example, the ratio of the wear ratings in Table II

$$\frac{R_A}{R_B} = \frac{92.6}{101.9} = 0.909$$

shows that tannage A was 9.1 per cent less resistant to wear than tannage B, whereas in the single comparison of the two tannages, the ratio 109.2 indicates that A was 9.2 per cent more resistant to wear than B. Thus an error of approximately 18 per cent would have resulted if the single comparison had been used to evaluate the relative wearing quality of the two leathers. In the same way, wear ratings show that tannages A and C differ by only 0.3 per cent in average wear resistance, whereas the 3 replicate ratios, 100.7, 114.0 and 72.4, indicate percentage differences of 0.7, 14.0 and —27.6, and the mean ratio, 94.0, indicates than tannage C wears 6.4 per cent better than tannage A. Fortunately, experimental errors may be decreased by closer control in small experiments, but even under such conditions, the variation in results makes it necessary to include a number of replicate comparisons in order to obtain reliable conclusions.

As the reliability of a method depends upon the agreement in the results of replicate tests, a comparison of results obtained by different methods should be of interest. In the experiment of Frey, Clarke and Leinbach[1] a vegetable tanned sole and a chrome retanned sole were used on each pair of shoes. One test, consisting of 18 replicate comparisons, resulted in ratios of thickness losses ranging from 1.00 to 1.93, whereas the rates of wear (hours per 9 irons) ranged from 94 to 839 for the vegetable tanned leather and from 137 to 1117 for the chrome retanned leather. The results show clearly that rates of wear vary much more than the corresponding ratios of thickness losses. Since wear ratios are influenced by only a few of the variables affecting rates of wear, the relative variance found in the experiment might have been expected.

Comparison with other methods brings out the following advantages of the method described here:

1. Wear ratings are affected only slightly by variations due to wearers and wearing conditions.
2. Wear ratings do not require measurements of wearing time, probably the least accurate and least consistent measurements in a service test.
3. Wear ratings are less variable and, therefore, more reliable than other ratings.
4. Since the new method is applicable in experiments with any number of leathers, it presents a standard method of testing and interpreting results.

Future developments in improving the quality of sole leather will require a satisfactory method of evaluation if they are to be given a fair chance of being accepted. Standardizing the method of testing and interpreting results should go a long way toward making such tests more readily understood and the conclusions more universally acceptable.

Summary

An improved method for interpreting data obtained in service tests on sole leather is described and is compared critically with methods formerly used. Since the method is applicable to experiments in which any number

of leathers are compared, it presents a standard method of testing and interpreting results. Its use is illustrated by calculations in data obtained in service tests of 20 tannages at Camp Lee, Va. Advantages of the method for future service tests are discussed.

Acknowledgment

The author is grateful to R. B. Hobbs of the National Bureau of Standards for furnishing the data used in the calculations and for helpful suggestions in the development of the method.

References

1. R. W. Frey, I. D. Clarke, and L. R. Leinbach, *J.A.L.C.A.*, **23**, 430 (1928).
2. L. M. Whitmore and G. V. Downing, *J.A.L.C.A.*, **37**, 150 (1942).
3. R. B. Hobbs and R. A. Kronstadt, *J.A.L.C.A.*, **40**, 12 (1945).
4. G. W. Snedecor, Statistical Methods. The Collegiate Press, Inc. (1940).